MULTI-MODALITY APPROACHES FOR MEDICAL SUPPORT SYSTEMS: WHERE WE ARE AND WHERE WE ARE GOING

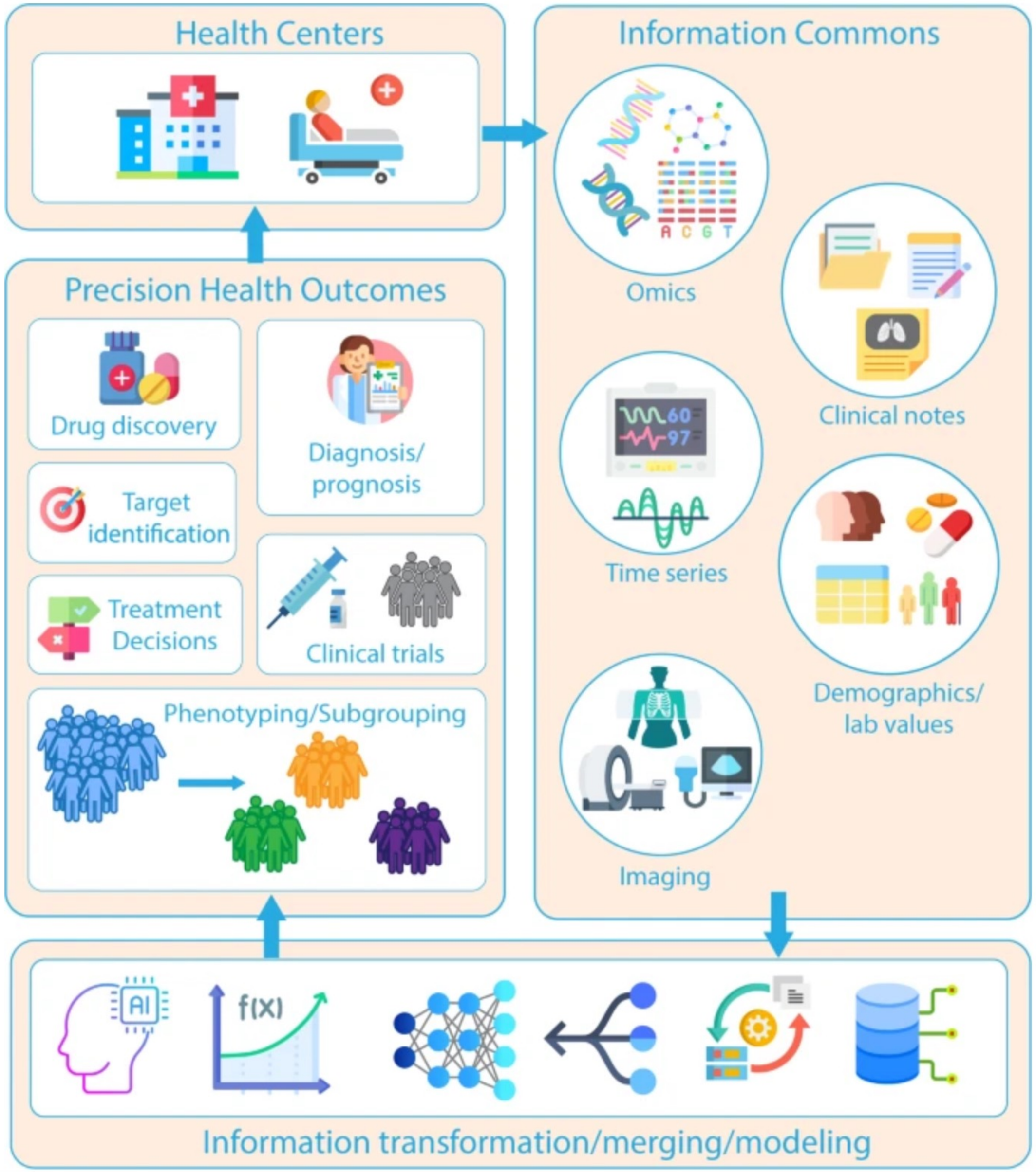
Salvador García <salvagl@decsai.ugr.es>

Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI)

University of Granada

SUMMARY OF THE TALK

- Limitations of traditional approaches:
  Typically rely on single-modality data, restricting medical decision-making.
- Technological advancements:
  Enable integration of diverse data sources for a more holistic patient view.
- What is a multi-modality approach?:
  Fusion and analysis of medical images, biosignals, clinical records, and more.
- Aim of this Speech:
  To explore multi-modality methods for disease diagnosis and prognosis.
- Relevance to personalized medicine:
  Allows detailed patient profiling including genetics, imaging, and clinical history.
- Technical challenges:
  Difficulties in fusing heterogeneous multimodal data.
- Emerging solutions:
  Deep learning as a powerful paradigm for multimodal data integration.



AGENDA OVERVIEW

- Introduction to Multi-Modality Approaches in Medicine
- Current Multi-Modality Technologies
- Advantages and Challenges of Multi-Modality Approaches
- Case Studies and Real-World Applications
- Current State of the Field: Where We Are
- Future Trends and Directions

INTRODUCTION TO MULTI-MODALITY APPROACHES IN MEDICINE

DEFINITION AND SIGNIFICANCE OF MULTI-MODALITY

- Integration of diverse data types:
  - Medical imaging, clinical text, sensors, genomics, and more
- Healthcare generates heterogeneous data:
  - EHR notes, radiology scans, lab results, wearable streams, genetic profiles
- Multimodal AI fuses data for deeper clinical insights
- Captures complementary information across modalities
- Enables:
  - More accurate diagnoses
  - Proactive patient monitoring
  - Personalized treatment strategies
- Example: Imaging + genomics = improved cancer diagnosis & planning
- Mimics physician reasoning, enhanced by AI pattern recognition

DEFINITION AND SIGNIFICANCE OF MULTI-MODALITY

Understanding Multi-Modality

Multi-modality involves integrating various methods and technologies to enhance data collection for informed decision-making. It involves fusing and analyzing various data types, including medical images, biosignals, clinical records, and other relevant sources.

Improving Diagnostic Accuracy

The combination of multiple modalities leads to improved accuracy in diagnoses, reducing the risk of errors.

For example, in Alzheimer's disease diagnosis, relying solely on structural MRI scans results in 80% detection accuracy.

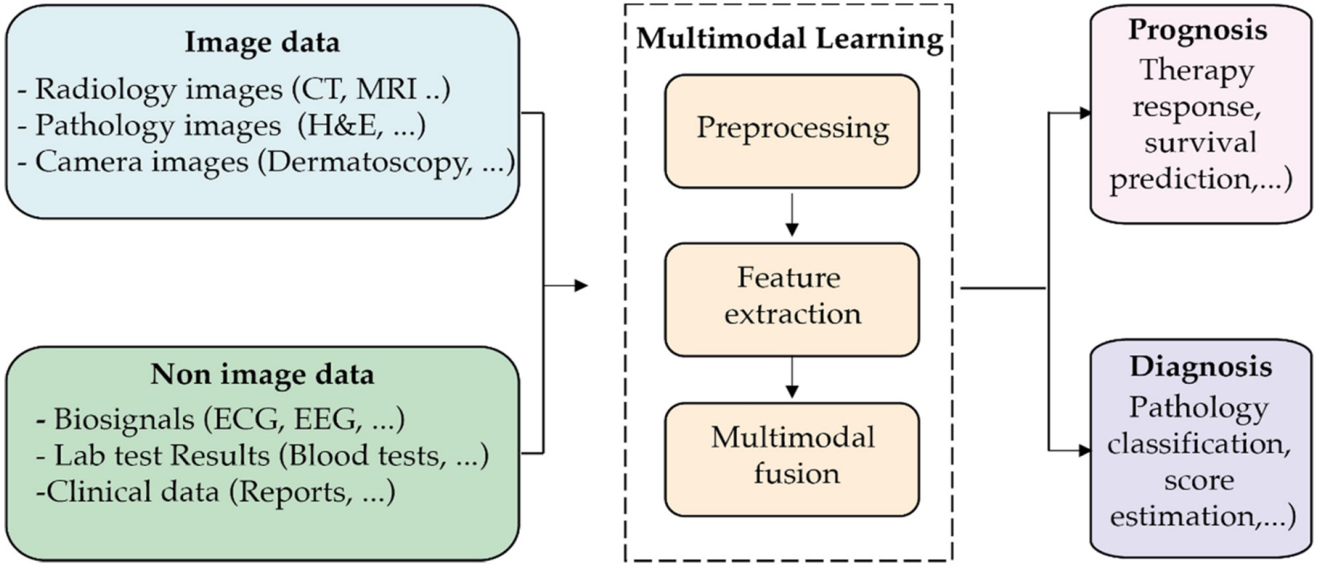
By also incorporating complementary modalities like audio features, speech transcript, genomic and clinical assessments, models have achieved over 90% diagnosis accuracy.

Personalizing Treatment Plans

Multi-modality allows for the personalization of treatment plans, catering to the unique needs of each patient.

DEFINITION AND SIGNIFICANCE OF MULTI-MODALITY

- Medical scanners are producing higher resolution digital images across modalities like MRI, CT, and PET.
- Electronic health records now compile diverse clinical data in structured formats.
- Advanced analytics methods like deep learning are capable of modeling complex multi-modal relationships [1].



HISTORICAL BACKGROUND AND EVOLUTION

Early medical diagnostics relied on isolated data types (e.g., physical exams, X-rays).

Technological milestones: invention of MRI (1970s), CT (1970s), PET scans (1980s).

Traditionally, each modality was analyzed separately, leading to fragmented clinical insights.

Growth of electronic health records (EHRs) enabled data integration.

Advances in computational power allowed fusion of complex modalities (images, signals, text).

Rise of machine learning and deep learning improved pattern recognition across modalities.

CURRENT MULTI-MODALITY TECHNOLOGIES

[2, 6]

IMAGING TECHNIQUES

MRI Technique

Magnetic Resonance Imaging (MRI) uses magnetic fields and radio waves to create detailed images of organs and tissues, providing crucial information for diagnosis.

CT Scanning

Computed Tomography (CT) scans combine X-ray images taken from different angles to produce cross-sectional views of bones and soft tissues, aiding in precise diagnosis.

PET Imaging

Positron Emission Tomography (PET) scans provide metabolic information by detecting radiation emitted from tracers injected into the body, assisting in the evaluation of conditions like cancer.

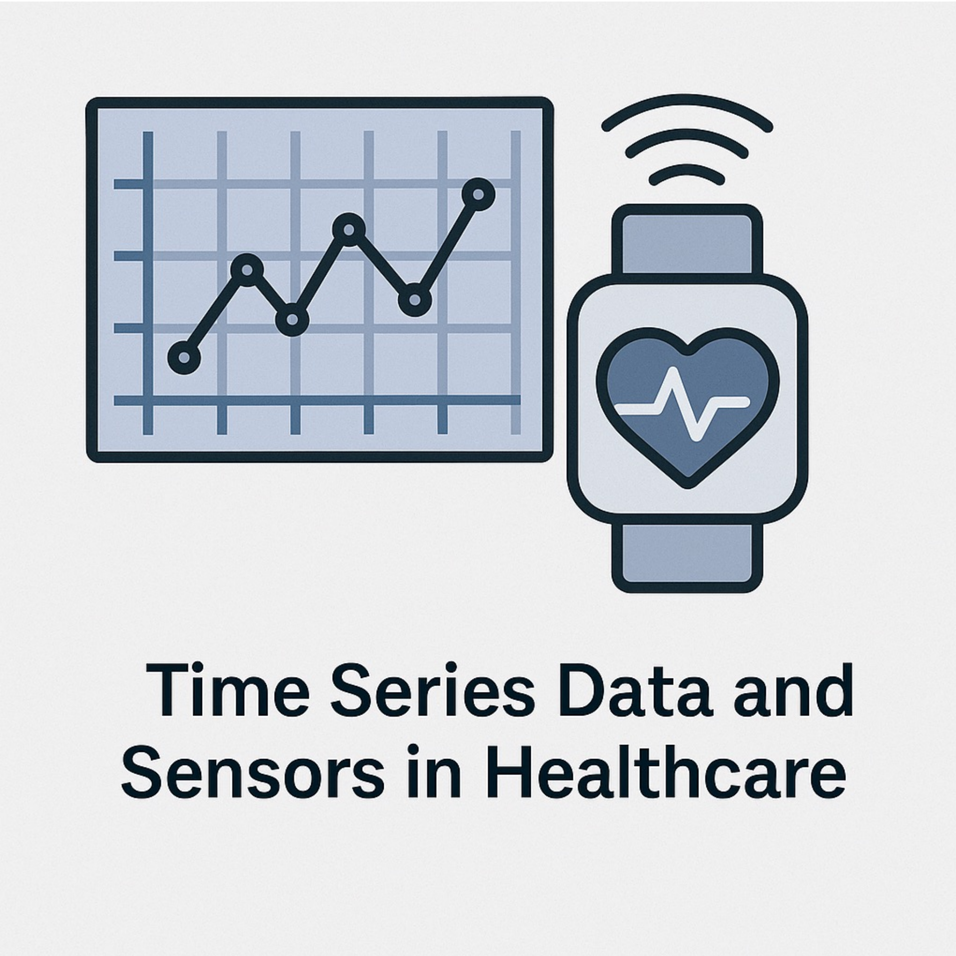
Imaging modalities such as X-rays, CT, MRI, ultrasound, and digital pathology slides provide detailed visual insights into anatomical structures and pathological conditions. These tools are essential for diagnostics, enabling the detection of abnormalities like tumors and monitoring disease progression or response to treatment.

Imaging offers unique spatial and structural information that complements textual and genomic data, enhancing decision-making and supporting personalized treatment strategies.

TEXTUAL AND ELECTRONIC HEALTH RECORDS (EHR) DATA

- **Clinical text** includes doctor's notes, radiology and pathology reports, discharge summaries, and other EHR documentation.
- Contains **rich contextual information** about patient history, symptoms, diagnoses, and treatments.
- Captures **nuances** not found in structured data, such as physician impressions and family history.
- **NLP techniques** extract key facts (e.g., diagnoses, medications) from unstructured narratives.
- In **multimodal applications**, textual data provides context to images or sensor readings (e.g., noting immunocompromised status for interpreting a chest X-ray).
- **EHR integration** improves predictive modeling; frameworks like COMET (2025) [7] use transfer learning from large EHR datasets to enhance smaller omics analyses.
- Adds **clinical semantics and patient-specific background**, enriching decision support beyond what images or signals alone provide.

SENSOR AND TIME-SERIES DATA

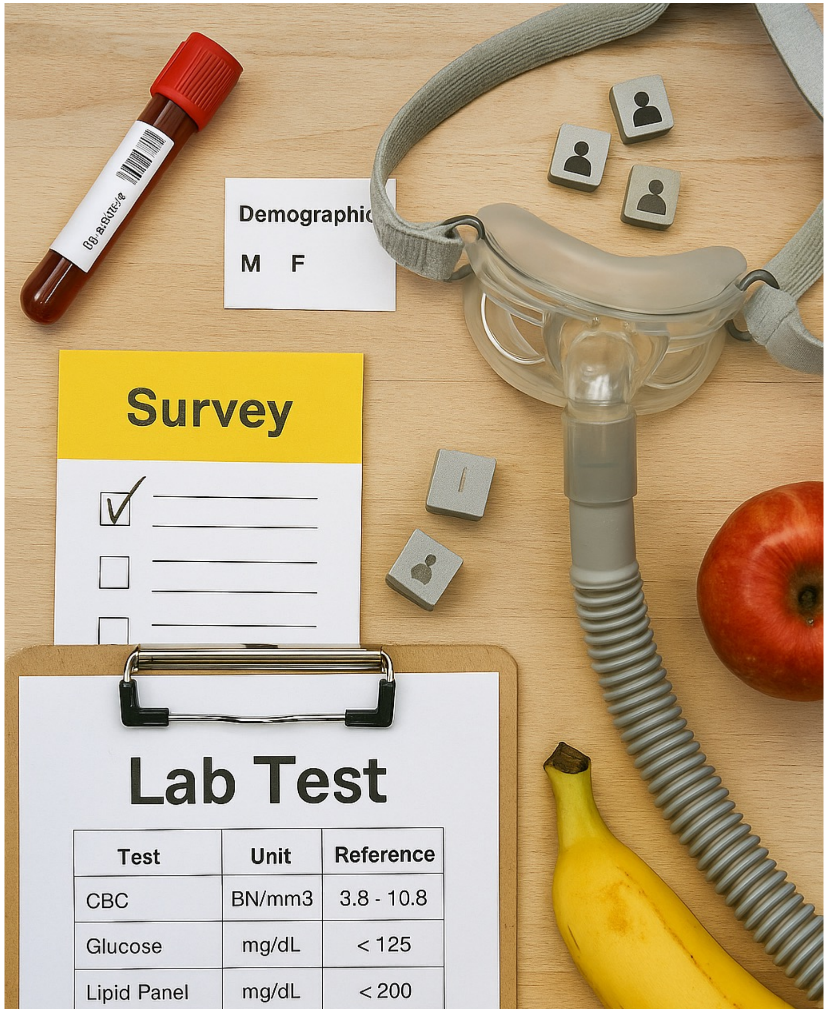


- **Wearable devices and hospital monitors** generate continuous physiological data (e.g., ECG, blood pressure, oxygen saturation, glucose levels, activity).
- Enable **real-time monitoring** and **early warning systems**, useful for remote care and chronic disease management.
- **Sensor data** supports **telemedicine** by alerting clinicians to patient condition changes remotely.
- In **multimodal systems**, sensor signals are combined with other data (e.g., symptoms, medications) for deeper context.
- Example: A rise in heart and respiratory rate may be interpreted differently when combined with recent clinical notes or lab results.
- Smart healthcare systems follow a **data lifecycle**: acquisition → structuring → fusion → predictive modeling.
- Sensors add the unique benefit of **continuous, real-time monitoring**, complementing traditional clinical measurements.

GENOMICS AND OMICS DATA

- **Genomic data** includes DNA sequences, genetic variants, gene expression, and molecular profiles (e.g., from tumor or germline sequencing).
- Provides insight into **biological mechanisms** of disease and drug response.
- Used to **diagnose hereditary conditions**, identify **genetic risk factors**, and guide **targeted therapies** (e.g., mutation-specific cancer treatments).
- **Decreasing sequencing costs** are making genomic profiling more accessible in routine care.
- **Integrating genomics** with imaging and clinical data supports more **precise, personalized medicine**.
- Adds a **molecular dimension** to multimodal systems, enhancing individualized predictions and decisions.
- When combined with **EHR and sensor data**, enables **comprehensive risk assessments** and **proactive monitoring** based on genetic predispositions.

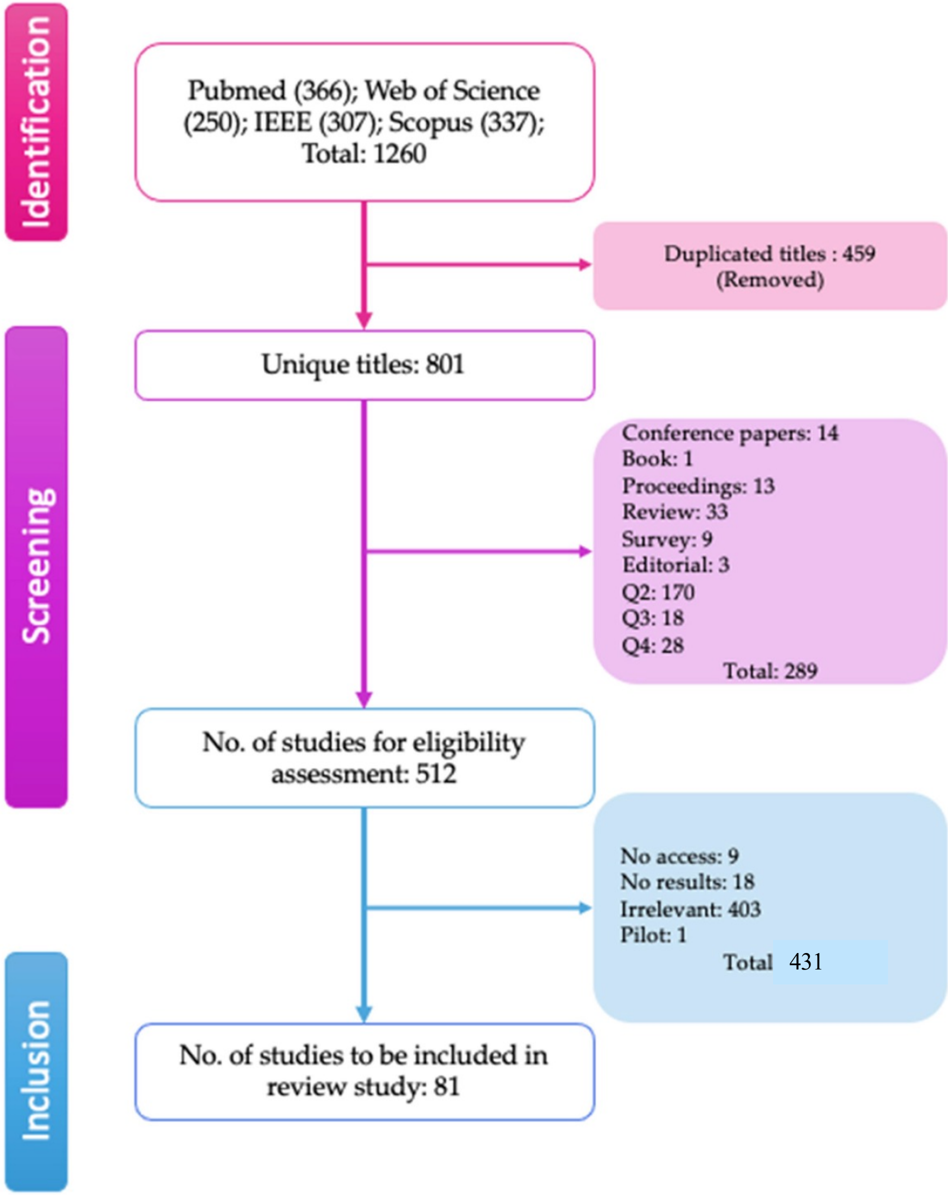
OTHER DATA MODALITIES



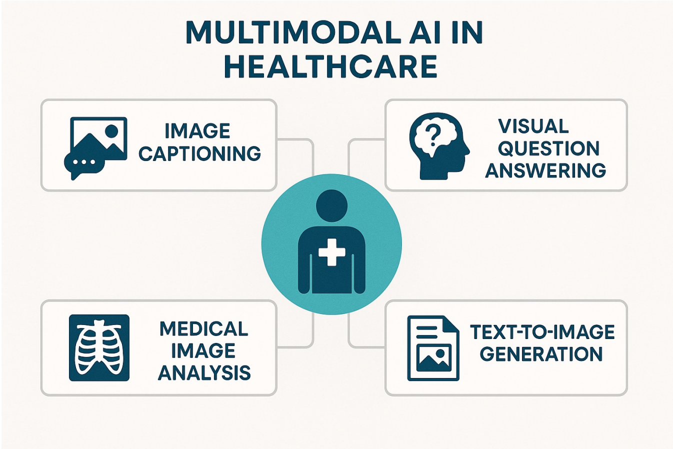
- **Laboratory tests** (e.g., blood tests, biomarkers) contribute structured numerical data.
- **Demographics** (age, sex, ethnicity) and **lifestyle data** (diet, exercise, sleep) inform overall health status.
- **Environmental data** (air quality, pollution, climate) aids in public health monitoring and identifying disease triggers.
- **Patient-reported outcomes** and surveys capture **subjective health indicators** like symptoms and quality of life.
- Multimodality may also include **social determinants of health** and **ambient data**.

SURVEY ON MULTI-MODALITY TECHNIQUES [1]

- Survey made in June 2023.
- 1260 articles reviewed (2012-2022).
- Only Q1 journals.
- Only multi-modals methods for detecting or predictin diseases.
- Machine learning or Deep Learning.
- Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines

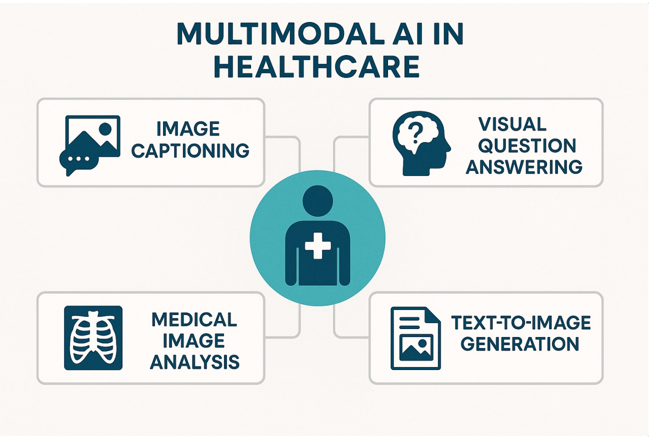


SURVEY ON MULTI-MODALITY TECHNIQUES: CATEGORIES



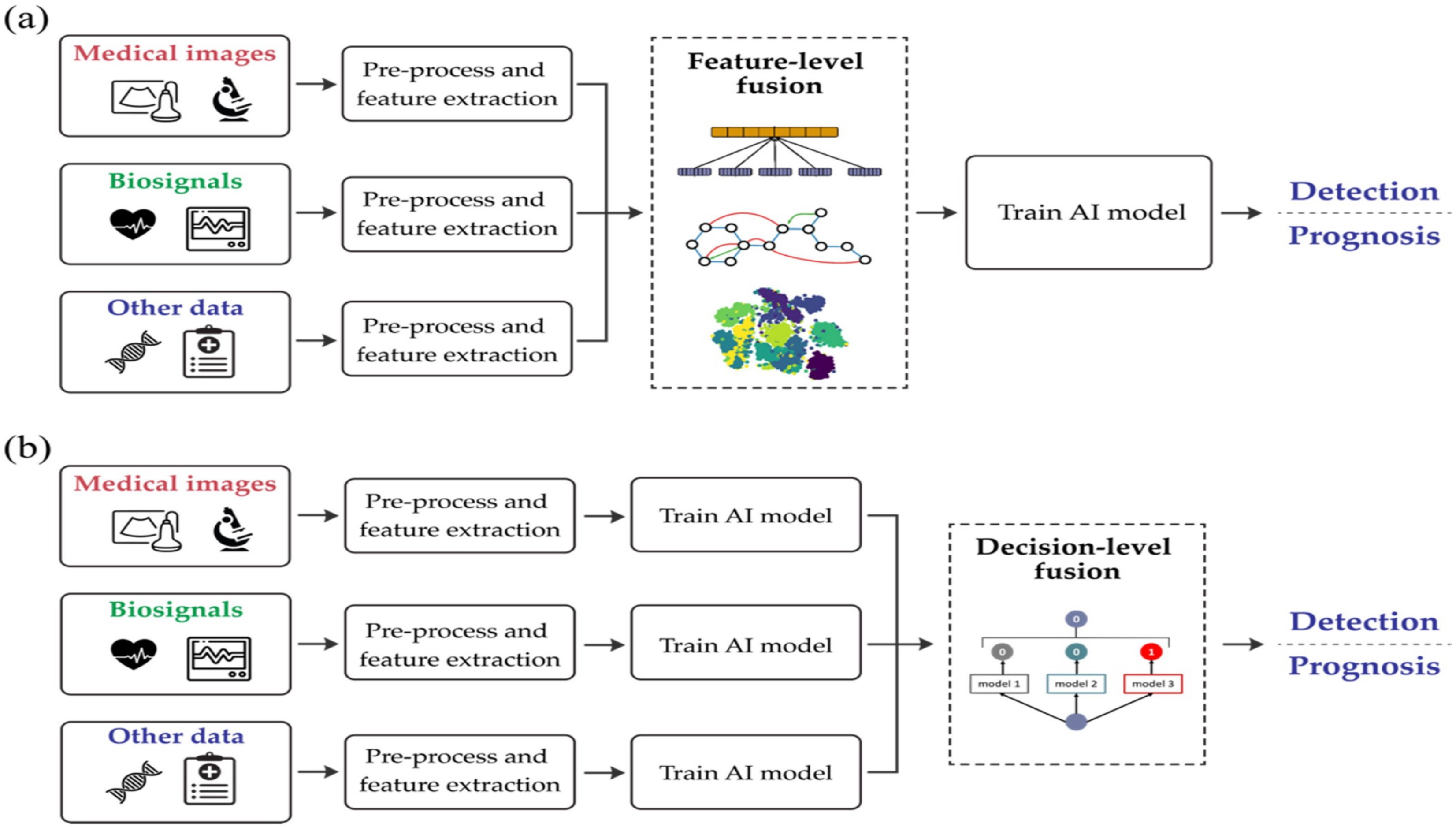
- Feature-level Fusion
  - Separate Feature Extraction: Features are extracted individually from each modality (e.g., MRI, blood tests), preserving their unique characteristics.
  - Fusion Strategy: Features are fused either:
    - Early Fusion: After initial processing but before classification.
    - Late Fusion: After complete modality-specific processing.
  - Joint Representation: The goal is to create a unified feature representation that combines all modality-specific features.
  - Complementary Information: This method leverages the strengths of different modalities to improve the quality of the final representation.
  - Use Case Example: MRI scans and blood tests are used to generate distinct feature sets, which are then combined for better classification performance.

SURVEY ON MULTI-MODALITY TECHNIQUES: CATEGORIES



- Decision-level Fusion
  - Independent Decision Making: Each modality (e.g., MRI, genetics) produces its own decision or prediction separately.
  - Combination Techniques: These individual decisions are fused using methods like:
    - Voting
    - Averaging
    - Weighted combinations
  - Example Use Case: MRI and genetics each classify a tumor as benign or malignant; their results are averaged for a final diagnosis.
  - Goal: Enhance decision-making by leveraging diverse, independent sources of information.
  - Advantages:
    - Allows flexibility in using different modalities.
    - No need for end-to-end joint training.
  - Limitation: Does not capture interactions between modalities as feature-level fusion does.

SURVEY ON MULTI-MODALITY TECHNIQUES:



CURRENT APPLICATIONS IN MEDICAL PRACTICE

In our study, three Main Categories of Data Types in Multi-Modality Approaches [1]:

Bioimaging:

Fusion of different medical imaging types.
Examples: PET + CT, PET + MRI (functional + anatomical images).

Biosignals:

Integration of physiological monitoring signals.
Examples: PPG (photoplethysmography), EEG (electroencephalography), ECG (electrocardiography).

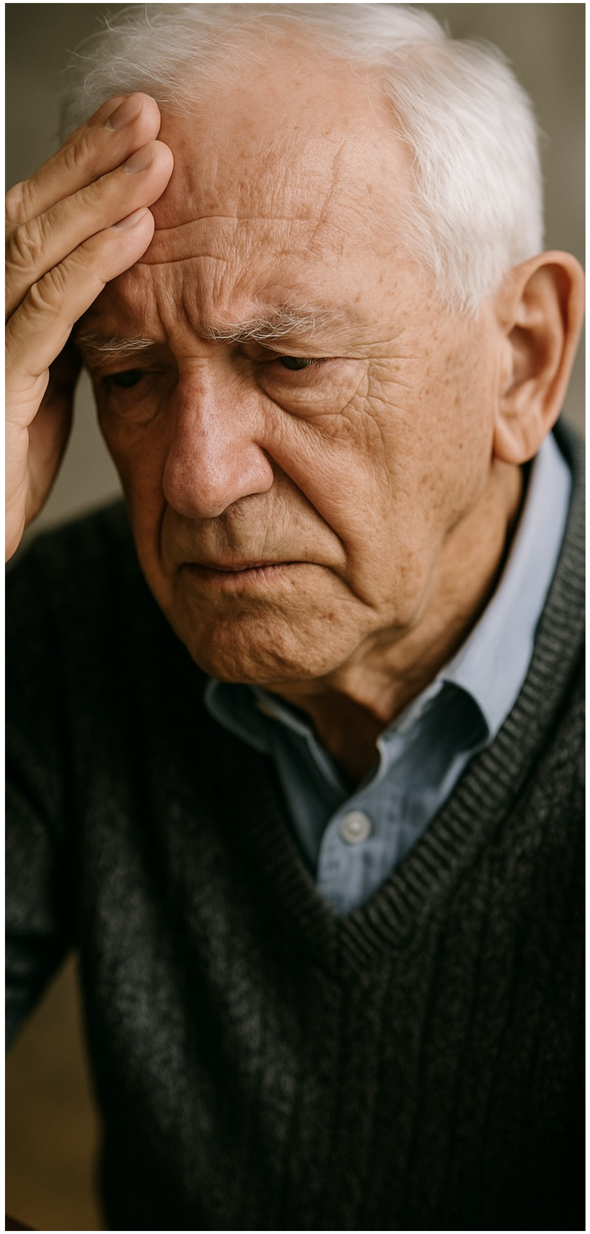
Mixed:

Combination of images, signals, and structured clinical data.
Examples: Medical images + biosignals + electronic health records (EHRs).

CLINICAL APPLICATION/DISEASE: COGNITIVE IMPAIRMENT

Multimodal AI integrates diverse data (e.g., imaging, genetics, clinical tests) to improve prediction and monitoring of cognitive decline in diseases like Alzheimer's and Parkinson's.

- One major application is Alzheimer's disease diagnosis and progression prediction using combinations of MRI, PET, and cognitive tests.
- Deep learning models (e.g., autoencoders, polynomial networks, LSTMs) improve classification by learning joint representations.
- Some studies combine longitudinal data, speech, language, and EEG for early detection and symptom tracking.
- Explainability techniques (e.g., SHAP, attention mechanisms) enhance trust in multimodal predictions.
- Smartphone sensors and genomic data are being explored for Parkinson's detection and risk prediction.
- **Challenges**:
  Combining diverse modalities, dealing with symptom variability, and ensuring interpretability remain key obstacles.



CLINICAL APPLICATION/DISEASE: MENTAL DISORDERS

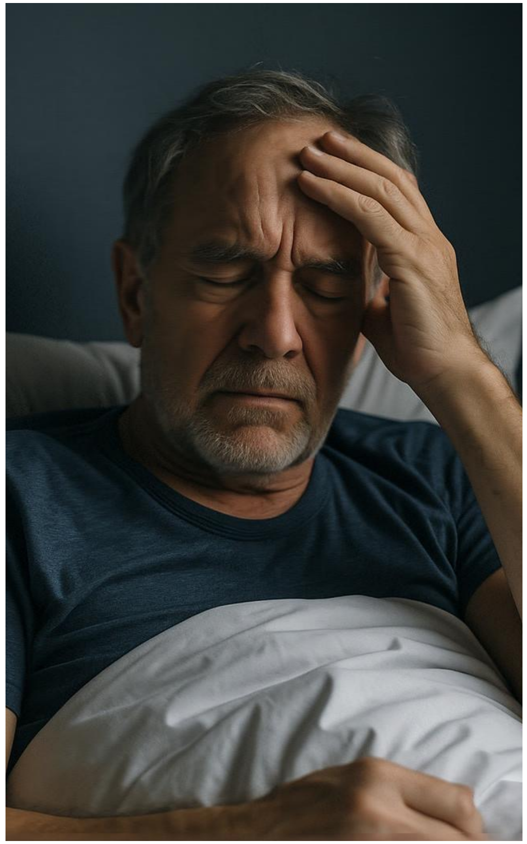
Multimodal AI combines imaging, biosignals, genetics, and clinical data to improve diagnosis, stratification, and treatment prediction in conditions like depression, schizophrenia, and autism.

- In depression, combining EEG, eye-tracking, and physiological data improves diagnosis and treatment response prediction.
- Autoencoders and graph networks have enhanced classification in autism and major depressive disorder by capturing joint behavioral and neural features.
- Schizophrenia prediction benefits from integrating structural/functional MRI with genomic data to uncover key neural and genetic factors.
- IoT-based and wearable biosensor systems enable real-time monitoring of stress and Parkinson's symptoms with low latency and energy efficiency.
- **Challenges**:
  High heterogeneity of data, complex correlations, and model optimization remain key obstacles for robust, generalizable mental health applications.

CLINICAL APPLICATION/DISEASE: SLEEP HEALTH

Multimodal AI integrates biosignals (EEG, EMG, EOG), clinical questionnaires, and wearable data to enhance sleep disorder diagnosis and monitoring beyond single-signal approaches.

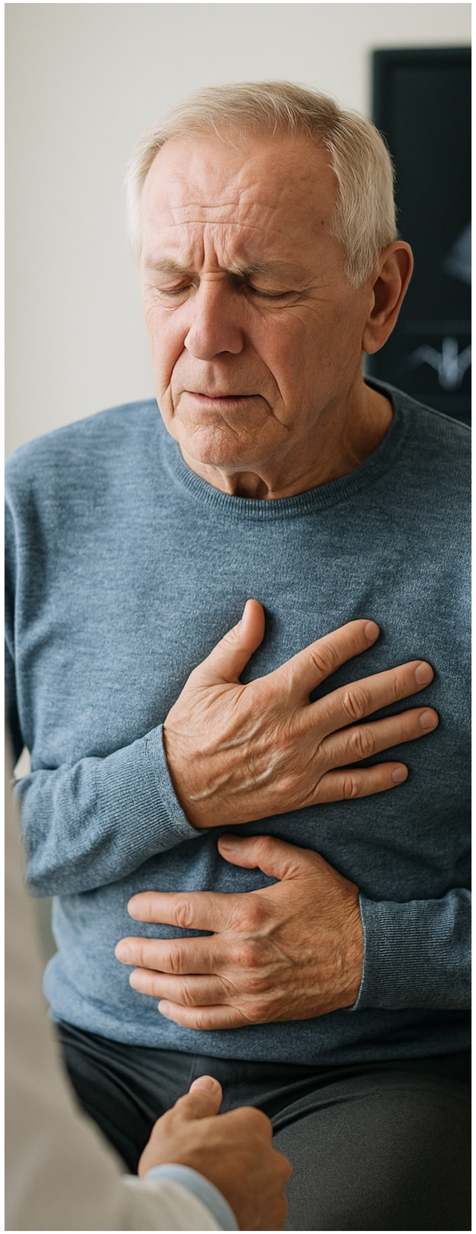
- LSTM models and smartphone/wearable data enable near-real-time sleep pattern detection with better performance than actigraphy.
- Fourier decomposition and multi-biosignal fusion (EEG, EMG, EOG) improve sleep stage classification over EEG alone.
- Models like SleepPrintNet and low-cost sensor setups demonstrate enhanced accuracy and potential for continuous patient monitoring.
- **Challenges**:
  Difficulties include extracting discriminative features across modalities, inter-subject variability, real-time computation limits, and modality-specific representation learning.



CLINICAL APPLICATION/DISEASE: CARDIAC DISEASES

Multimodal AI integrates ECG, imaging (MRI, ultrasound), EHRs, and genetics to enhance diagnosis, risk assessment, and treatment planning in cardiac diseases like heart failure and arrhythmias.

- Combining 3D MRI and ultrasound motion data with machine learning improved classification of dilated cardiomyopathy.
- Feature fusion and hybrid selection boosted coronary artery disease detection accuracy.
- Integrative systems using EHR and physiological signals predicted postoperative cardiac events with high performance.
- **Challenges**:
  Key issues include fusing heterogeneous data types, generalizing across diverse patient populations, and managing the complex etiology of cardiac conditions.



CLINICAL APPLICATION/DISEASE: COVID-19

Multimodal AI integrates clinical data, lab tests, imaging, audio, and outcomes to predict COVID-19 severity, track progression, and guide treatment.

- Random forest models using clinical and lab data accurately differentiate severe vs. non-severe cases.
- CovScanNet fused breathing sounds and chest X-rays, achieving high accuracy via smartphone-based screening.
- Audio (cough/breathing) and imaging fusion improved early detection, while knowledge graphs enhanced doctor-patient communication.
- **Challenges**:
  Integrating diverse data types (text, audio, images), ensuring real-time performance, and handling multi-center data standardization and privacy are major hurdles.

CLINICAL APPLICATION/DISEASE: ONCOLOGY

Multimodal AI fuses imaging, histopathology, omics, and clinical data to improve cancer detection, subtype classification, treatment planning, and survival prediction.

- Deep learning models integrating MRI, pathology, and clinical data have improved diagnosis and recurrence prediction in prostate, breast, and brain cancers.
- Multimodal methods enhance glioma segmentation, subtype classification, and gene mutation prediction from MRI and intraoperative data.
- Applications also include thyroid, pancreatic, liver, gastric, and cervical tumors using ultrasound, PET-CT, and CT combined with clinical records.
- **Challenges**:
  Fusion of diverse modalities, handling 3D imaging in 2D models, increased computational demand, and integrating domain knowledge remain key obstacles.

CLINICAL APPLICATION/DISEASE: OPHTHALMOLOGY

Multimodal AI integrates retinal imaging (fundus, OCT), clinical data, and text to improve diagnosis and management of eye diseases like macular degeneration, glaucoma, and retinitis pigmentosa.

- Bayesian and self-supervised models improved glaucoma detection by combining imaging, indicators, and clinical text while accounting for uncertainty.
- Deep learning with fundus and OCT images enhanced detection of choroidal neovascularization and prediction of visual impairment.
- Multimodal pre-training strategies improved grading accuracy across datasets using unlabeled retinal image pairs.
- **Challenges**:
  Capturing interactions across modalities, managing diagnostic uncertainty, and exploiting unlabeled data for robust learning are key hurdles in ophthalmic AI.

CLINICAL APPLICATION/DISEASE: PEDIATRIC DISORDERS

Multimodal AI combines imaging, clinical data, biosignals, and behavior to improve early detection, intervention planning, and monitoring of pediatric conditions.

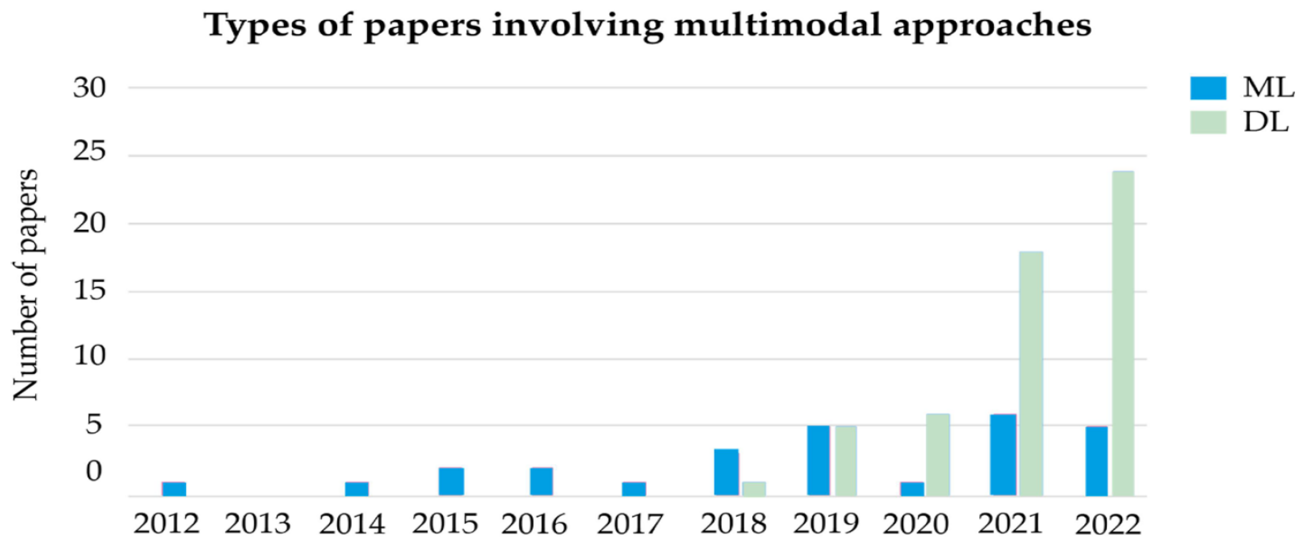
- MCNNs using fetal heart rate and contractions aided fetal compromise prediction, highlighting the need for hybrid clinical-AI models.
- Radiograph + clinical data fusion improved diagnosis and surgery prediction in necrotizing enterocolitis.
- Video, audio, and physiological signals enabled accurate neonatal pain assessment; imaging + biomarkers enhanced Crohn's disease activity prediction.
- **Challenges**:
  Limited pediatric data, high imaging costs, privacy concerns, and lack of age-standardized tools hinder scalable and reliable multimodal integration.

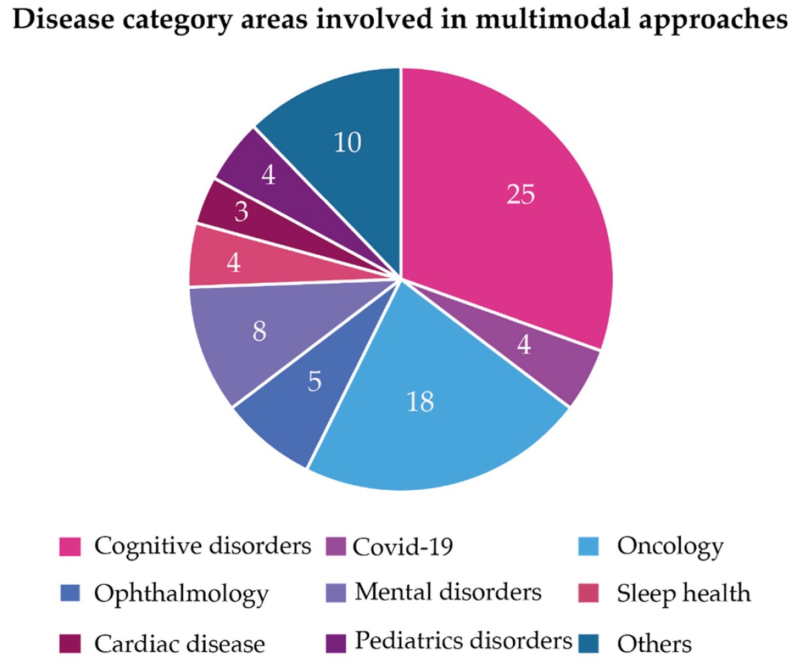
CLINICAL APPLICATION/DISEASE:
OTHER STUDIES/MISCELLANEOUS APPLICATIONS

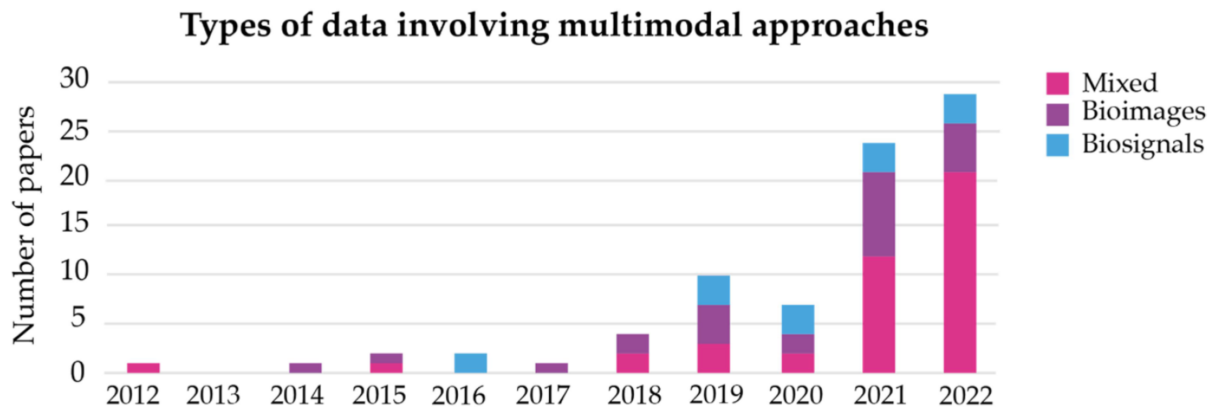
These diverse studies apply multimodal AI to improve detection, diagnosis, and classification across a range of conditions using combined clinical, imaging, and physiological data.

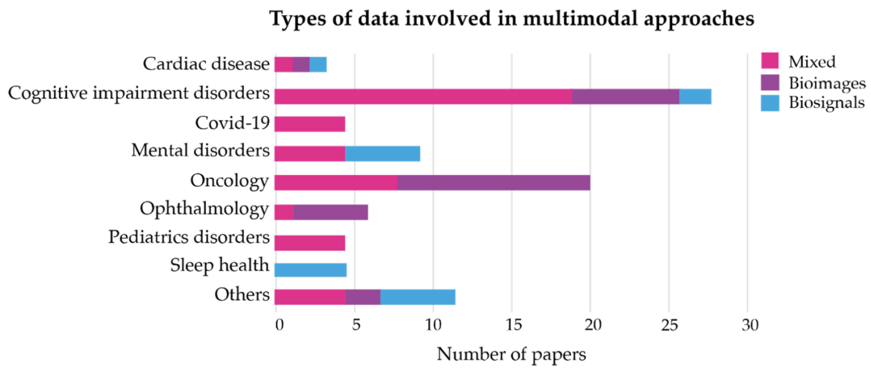
- Seizure detection with EMG and accelerometry improved recognition of short, nonstandard events.
- Multimodal models enhanced infection screening, lung nodule and liver fibrosis staging, and knee osteoarthritis prediction.
- Combining CT, ultrasound, PET/CT, and clinical data improved diagnosis in pulmonary embolism, cervical cancer, and lymph node metastasis.
- **Challenges**:
  Heterogeneity of conditions and modalities, limited standardized datasets, and integration of structured/unstructured data complicate generalization and deployment.

SUMMARY OF MAIN FINDINGS



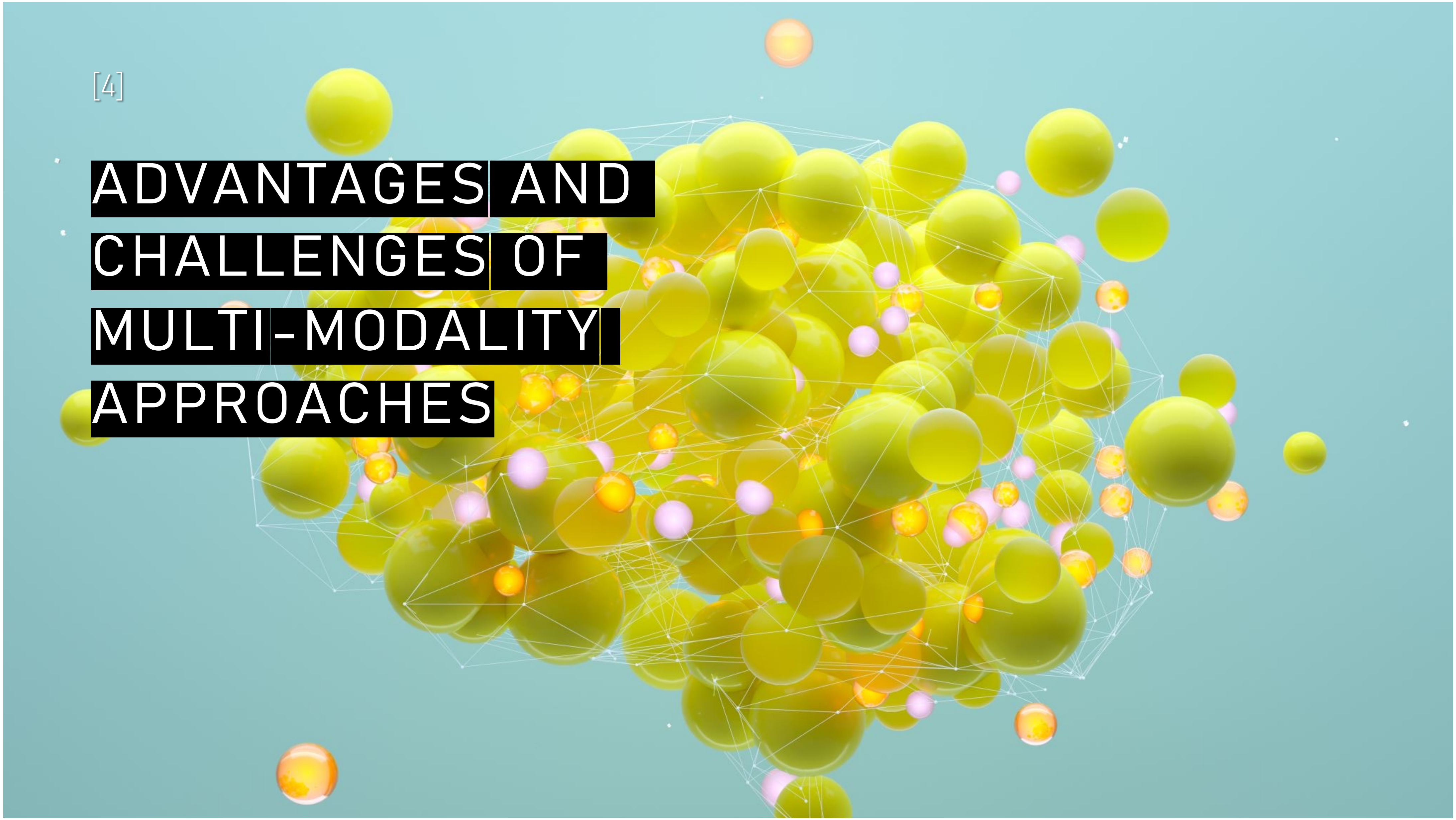






[4]

ADVANTAGES AND CHALLENGES OF MULTI-MODALITY APPROACHES



IMPROVED DIAGNOSTIC ACCURACY AND PRECISION

Real-World Impact Example [9]:
Breast cancer study showed improved diagnostic accuracy by combining mammogram + MRI
→ Accuracy increased from 93% (MRI only) to 99% (multimodal)
→ Enabled earlier treatment and improved patient outcomes

Key Benefits of Multimodal Integration:

✅ Improved diagnostic accuracy
→ Combines complementary data for better disease detection

🔍 More reliable detection
→ Captures subtleties missed by single tests

🧠 Deeper understanding of disease
→ Reveals mechanisms, stage, and heterogeneity

🧬 Supports personalized treatment
→ Enables better selection of targeted therapies

IMPROVED PATIENT OUTCOMES AND PERSONALIZED TREATMENT

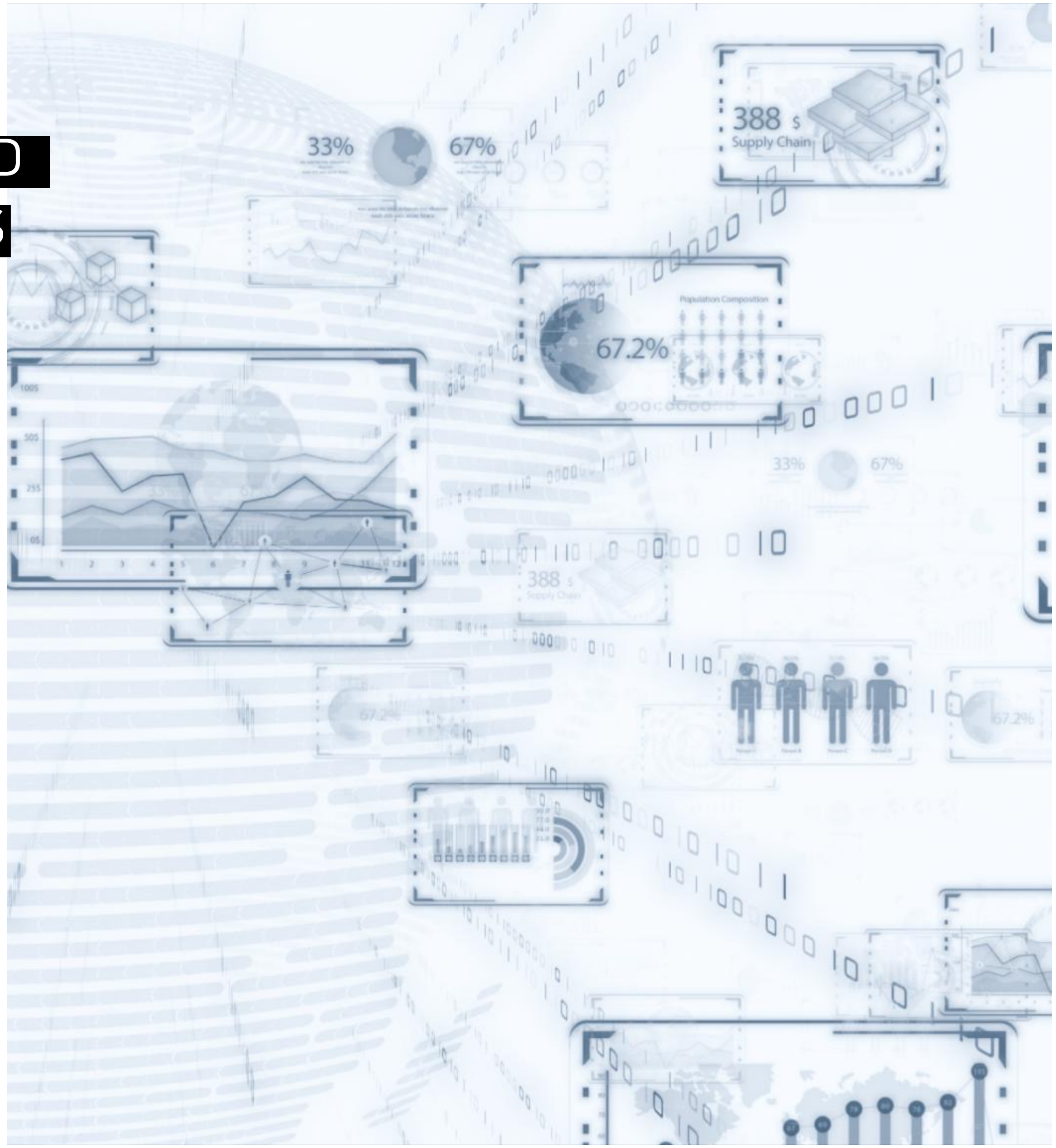
🧩 Holistic patient view
→ Combines genetics, lab results, clinical history, and imaging
→ Understands the patient **as a whole**, not just the disease

🧠 **Predictive insights for treatment**
→ Identifies optimal therapy paths and likely treatment responses

💡 Personalized, data-driven care
→ Tailors interventions to individual needs for better outcomes

❤️ Focus on quality of life
→ Empowers **patient-centered** decisions and improved long-term well-being

CHALLENGUES: TECHNICAL, LOGISTICAL, AND COST-RELATED CHALLENGES



Technical Complexities

Multi-modality approaches often involve sophisticated technologies that can be difficult to implement and manage.

Logistical Coordination

Coordinating different modalities requires careful logistical planning and management to ensure smooth operation.

Cost Implications

The advanced technologies involved in multi-modality approaches can lead to significant cost challenges for organizations.

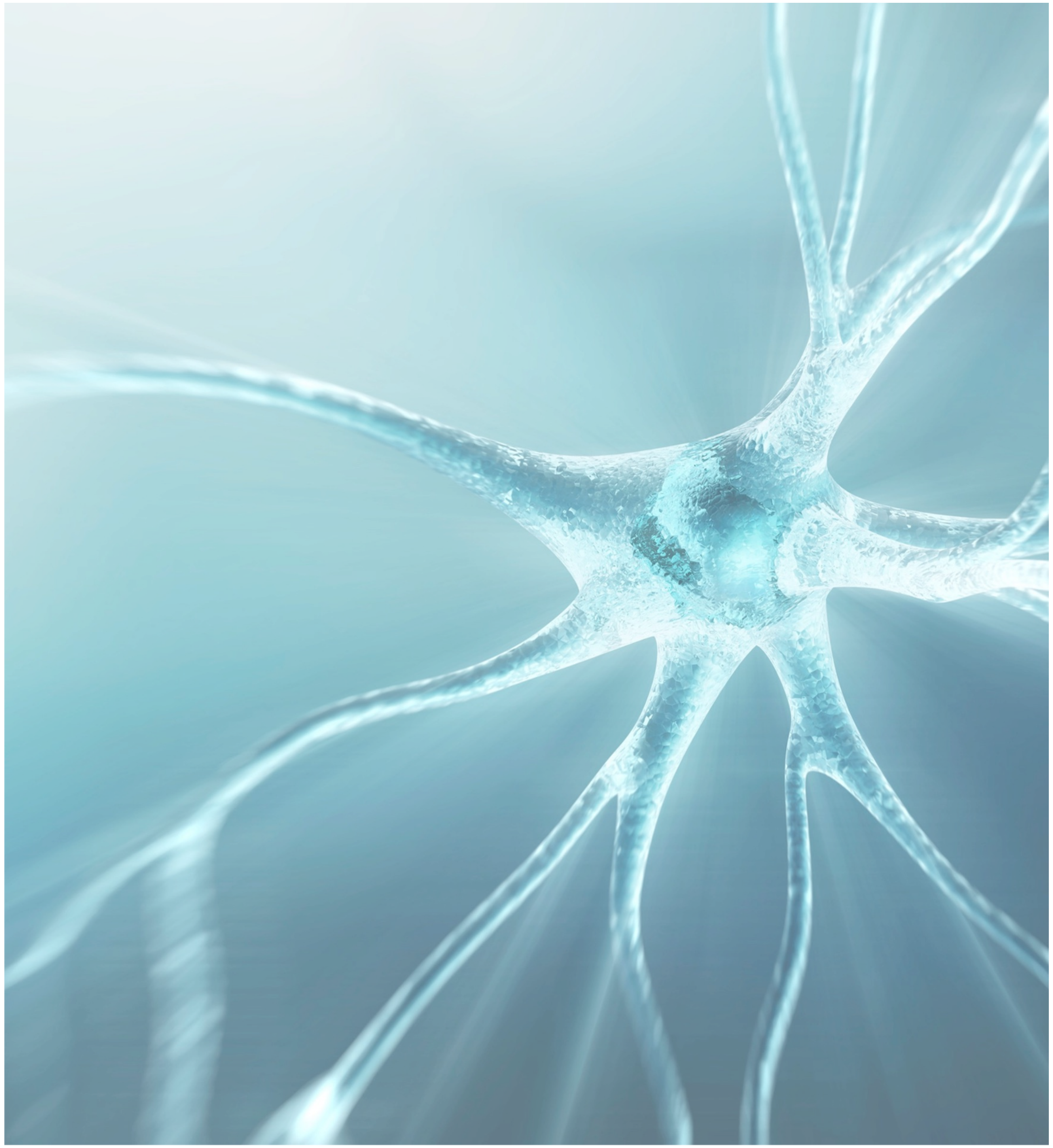
CHALLENGUES

Reliable and Time-Efficient Data Fusion

- Current fusion techniques may lose important information
- In some cases, single modalities outperform multimodal models
- Limited performance gains in certain studies highlight the need for better fusion strategies
- High computational costs hinder real-time application in clinical settings

Balancing Model Complexity with Interpretability and Transparency

- Multimodal models tend to be complex and less interpretable
- Interpretability is essential for clinical adoption and trust
- Requires new methods to enforce transparency or simplify model mechanics
- Essential to design models that justify decisions for critical healthcare applications



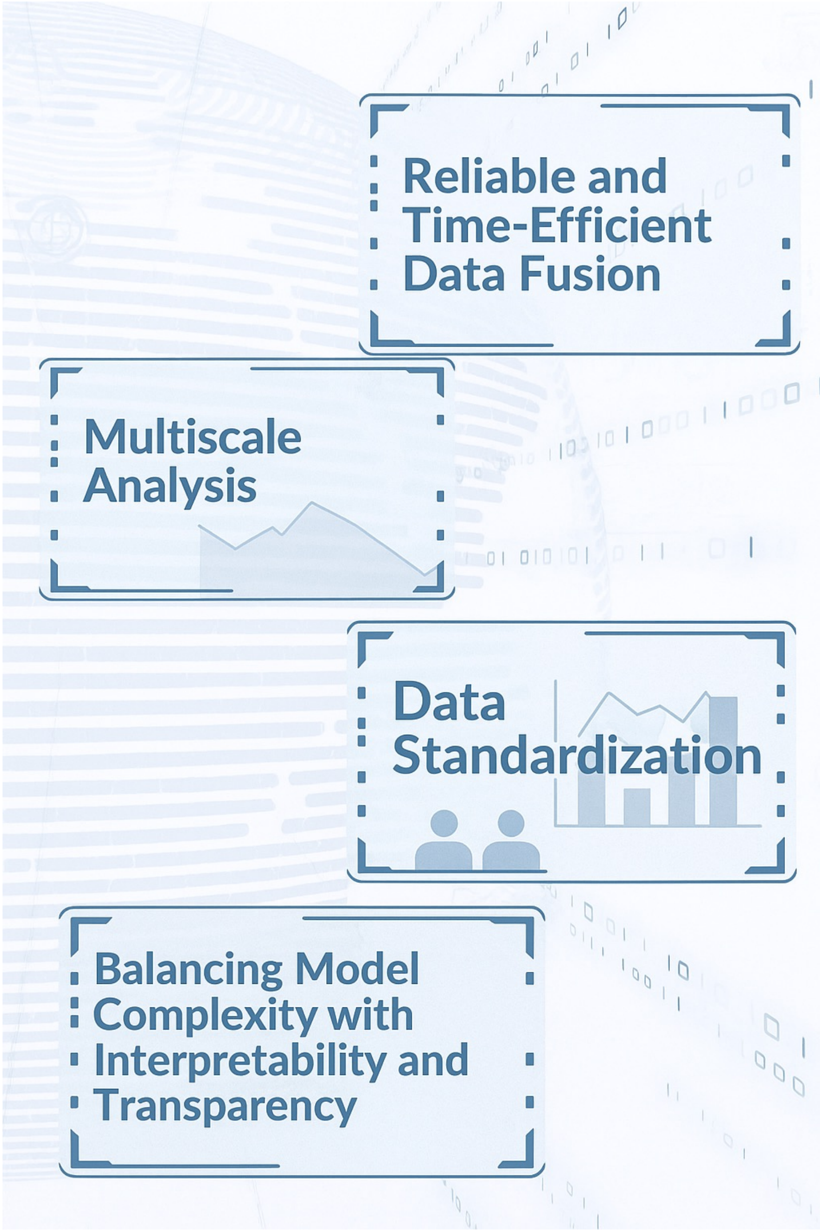
CHALLENGUES

Multiscale Analysis

- Modalities capture information at different biological scales (e.g., molecular vs. organ-level)
- Genetics and proteomics provide fine-grained data; imaging reveals anatomy and function
- Integrating data with different spatial and semantic resolutions is complex and remains a major challenge

Data Standardization

- Variability across centers, scanners, and acquisition protocols can introduce bias
- Harmonization and normalization are essential before data fusion
- Standardization pipelines must preserve true biological signals while minimizing technical noise



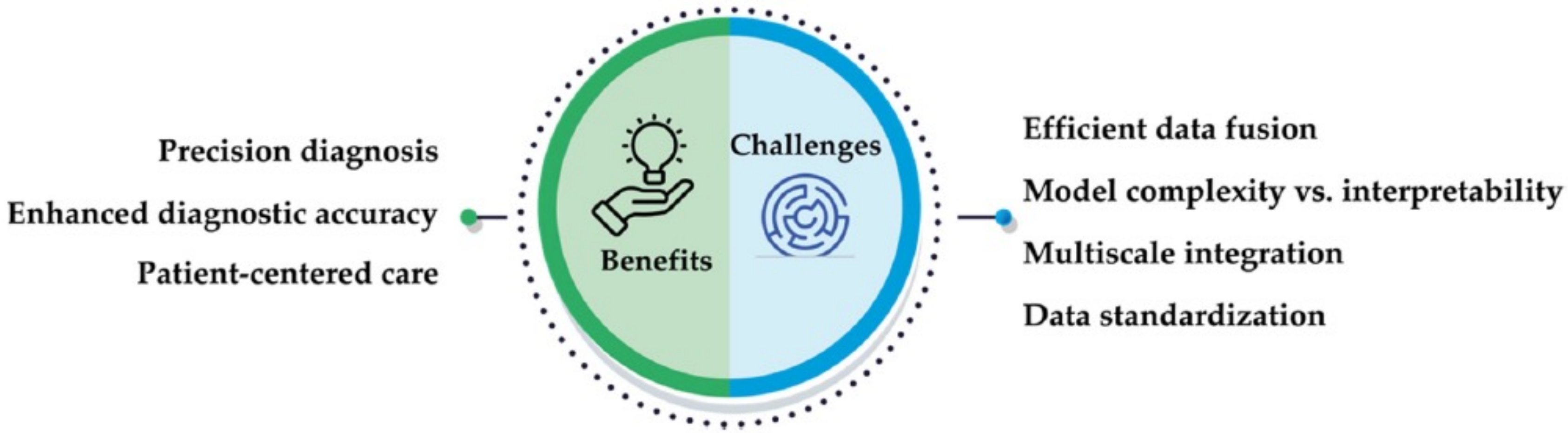
ADDRESSING CHALLENGUES IN MULTIMODAL HEALTHCARE

Technical Approaches

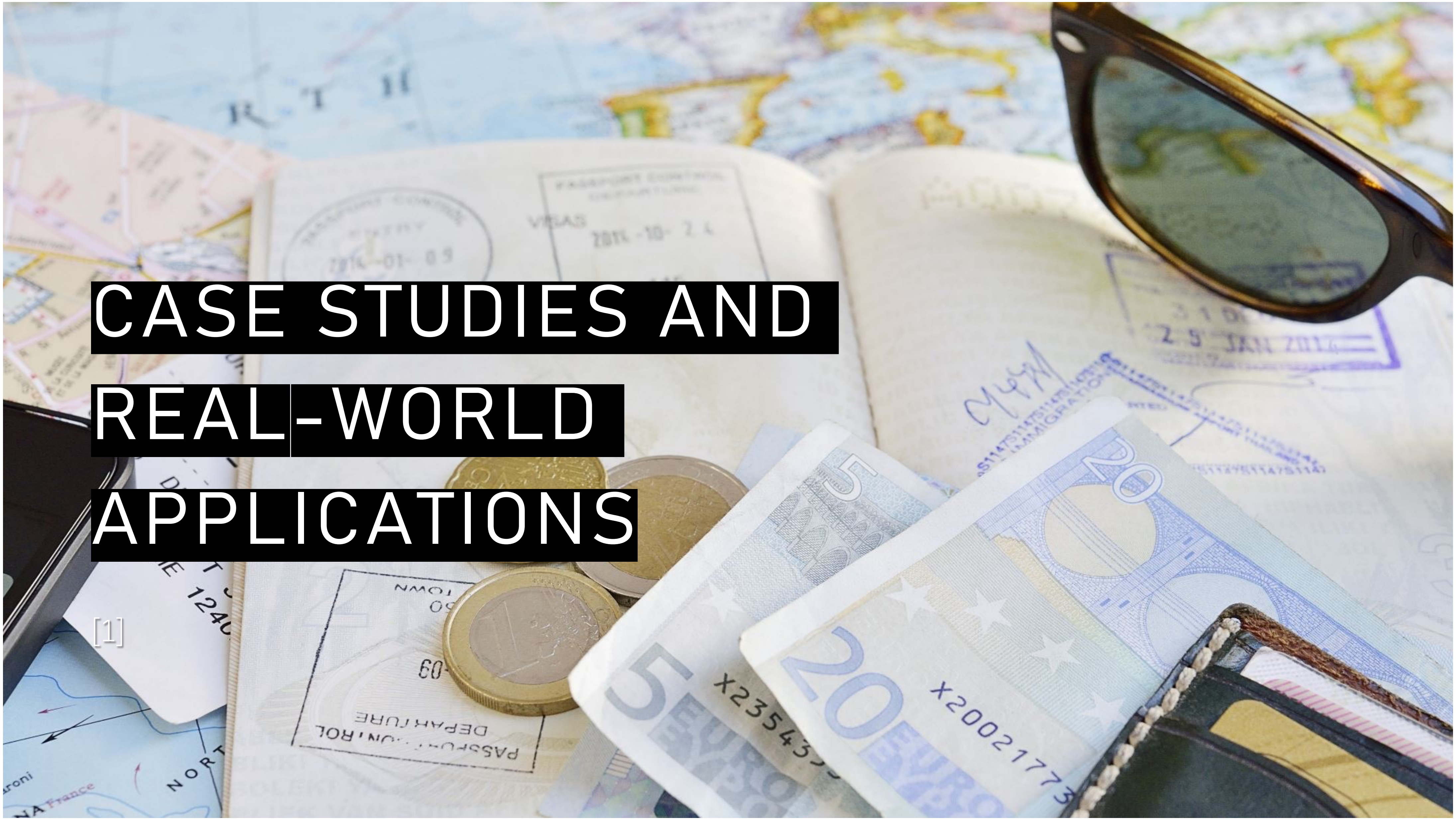
- *Multi-view learning*: Exploits complementary perspectives from different data modalities
- *Adversarial training*: Improves robustness and generalization across modalities
- *Attention mechanisms*: Enhances focus on the most relevant features across diverse inputs
- Despite advances, **significant progress** is still needed for effective integration of heterogeneous data

Ethical Considerations

- *Patient privacy and data security* must be prioritized
- *Informed consent* is essential for ethical data use
- *Data anonymization* and *ethical approvals* are required in all studies
- Establish **robust governance frameworks** and consent protocols to ensure trust and responsible implementation



CASE STUDIES AND REAL-WORLD APPLICATIONS



[1]

MULTIMODAL APPLICATIONS IN DIAGNOSTICS

- Enhanced Disease Detection with Multimodal AI
  - Multimodal models improve disease prediction.
  - Review of 128 studies (2011–2021): average **+6.4% accuracy** vs. single-modality models [6]
  - **Oncology**: Radiology + pathology + genomics = better cancer detection and subtyping.
  - **Neurology**: MRI/PET + clinical/genetic data enhance Alzheimer's and epilepsy diagnosis.
  - Forbes Tech Council [3]: Multimodal fusion gives a "clearer, more complete picture" for treatment planning.

- Contextualizing Imaging with Clinical Data
  - Radiology models boosted by EHR context (fever, surgery history, immune status, etc.).
  - Chest X-ray + patient data = improved pneumonia detection, fewer false positives.
  - **Ophthalmology**: Retinal scans + systemic data predict cardiovascular risks not seen in the eye.
  - **Explanatory report generation**

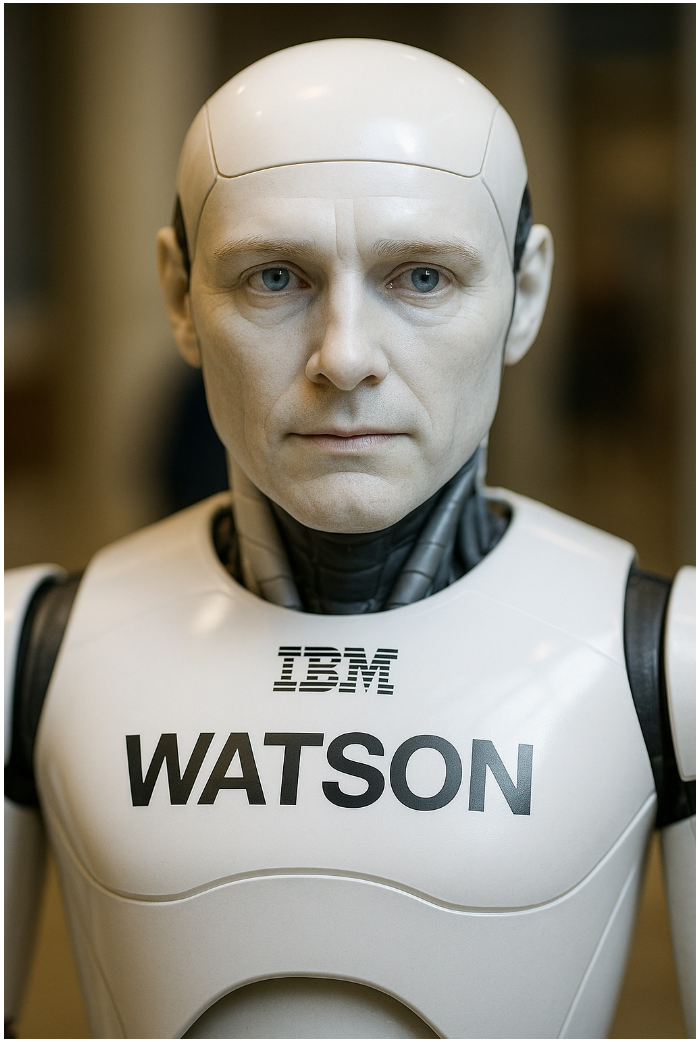
- Multi-Domain Diagnoses with Multimodal AI
  - Some conditions (e.g., genetic diseases) require **genomics + clinical data** for accurate diagnosis.
  - AI can detect **phenotypic patterns** (EHR, labs) and suggest genomic testing — or vice versa.
  - **Infectious diseases**: Accurate diagnosis may need **CT images + lab results + travel history**.
  - Fusing data types helps AI systems **replicate specialist reasoning** for complex cases.

SUCCESS STORIES IN ONCOLOGY DIAGNOSTICS

- **Tumor boards** use AI to integrate pathology slides, radiology, and molecular data.
- Models combining **pathology + genomics** improve tumor **subtype and grade classification**.
- **Breast cancer**: Mammography + patient risk factors (family history, biopsies, genetics) = tailored screening.
- **Oncology & neurology** dominate multimodal diagnostic research by volume.
- Multimodal data capture cancer’s **heterogeneity**, forming a rich diagnostic “signature”.

EXAMPLE – IBM WATSON FOR ONCOLOGY

- IBM Watson integrated **EHR text, labs, guidelines**, and later **genomic data** for cancer treatment suggestions.
- Sought to match patient data with medical literature to **recommend therapies**.
- Despite its ambition, Watson **often disagreed with expert oncologists**.
- In *Convergence of evolving artificial intelligence and machine learning techniques in precision oncology (2025)*: "Failed to achieve high concordance with expert clinicians."
- **Key takeaway**: Multimodal AI needs **rigorous validation** and close alignment with clinical workflows.
- Still, Watson showcased the **potential of AI to synthesize diverse medical data** for decision support.



MULTIMODAL APPLICATIONS IN PROGNOSTICS

- Outcome Prediction (EHR + Clinical Notes)
  - Combining structured clinical data (e.g., vitals, labs) with unstructured physician notes boosts predictive accuracy.
  - 2024 study: Multimodal model for in-hospital heart failure mortality → AUC ~0.84.
  - Textual notes added symptom severity, comorbidities, and illness history.
  - Outperformed unimodal baselines (structured-only or text-only) across all test sets.

- Survival & Progression Forecasting
  - Oncology: Integrates pathology images, genomics, and clinical staging to predict 5-year survival or therapy response.
  - The Cancer Genome Atlas (TCGA)-based models outperform single-modality ones [2].
  - Alzheimer's: MRI, biomarkers, cognitive scores, and APOE genotype improve dementia progression prediction.
  - ADNI dataset is a benchmark for multimodal Alzheimer's prognosis [2].

https://www.jmir.org/2024/1/e54363/

MULTIMODAL APPLICATIONS IN PROGNOSTICS

Treatment Outcome Prediction

- Cardiology: Uses imaging (e.g., echo), sensors (e.g., ECG), and clinical variables post-surgery.
- Oncology: Multi-omics + EHR predict therapy efficacy and risk of side effects [7].
- COMET framework (2025): Pretrained on EHR, fused with omics data for better small-cohort predictions.
- Enables precise patient subgrouping beyond standard case/control.

Remote Monitoring & Early Warning

- Multimodal early warning systems combine time-series vitals, labs, and nursing notes.
- ICU: Fusing monitor data + labs + bedside notes forecasts deterioration (e.g., septic shock).
- Traumatic brain injury: Neuroimaging + EEG + notes improved neurological recovery prediction.

Population-Level/Public Health Prognostics

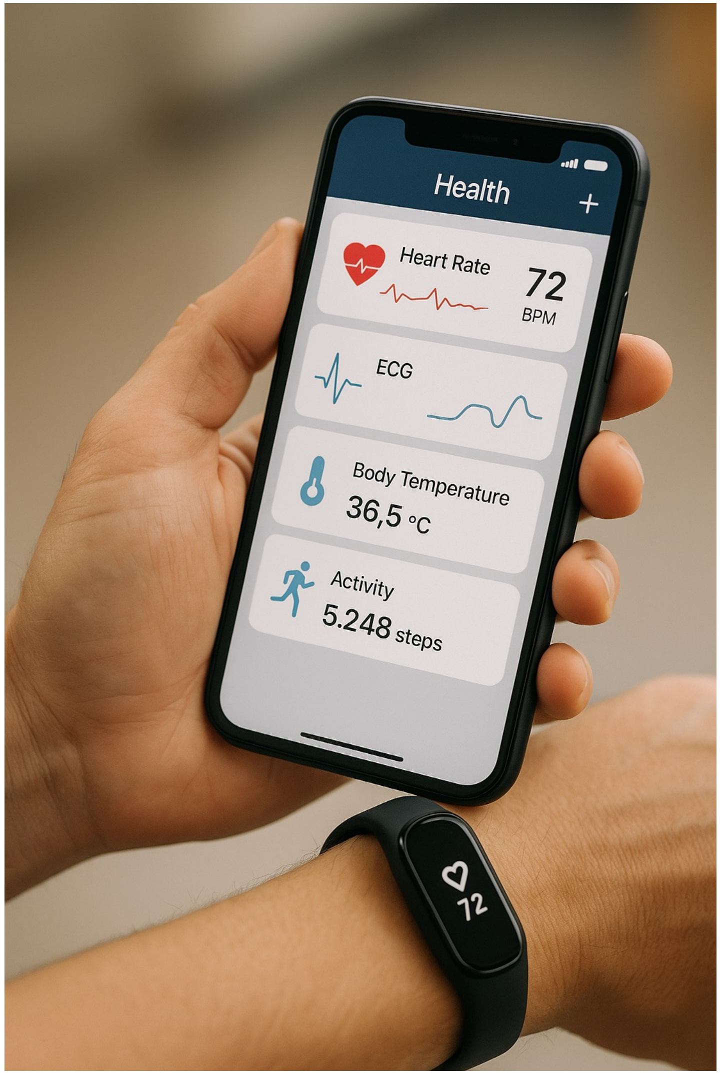
- Combines epidemiology, social media, environmental sensors, and pathogen genomics.
- COVID-19 modeling: Clinical data + mobility sensors + viral genome to predict outbreaks.
- Demonstrates scalability and versatility of multimodal predictive systems.

MULTIMODAL APPLICATIONS IN PATIENT MONITORING

- Remote Monitoring & Telehealth
  - Rise of wearables & IoT enables continuous out-of-hospital monitoring
  - Platforms integrate:
    - Physiological data (heart rate, temp, etc.)
    - Self-reported symptoms via apps
    - Medication adherence & telehealth assessments
  - Example: **Biofourmis Biovitals® Sentinel** (COVID-19, Singapore)
    - Clinical-grade biosensors + symptom app + AI analytics
    - Early alerts reduced in-person checks & protected staff
  - "Hospital at home" kits combine pulse oximeter, thermometer
  - Multimodal fusion (sensor trends + symptoms) yields better early warnings

- Intensive Care & Inpatient Monitoring
  - ICU patients produce high-frequency, multimodal data (ECG, labs, imaging)
  - AI fuses vital signs + labs + notes to predict events (e.g., septic shock)
  - Techniques: temporal convolution, attention models for real-time synthesis
  - Multimodal correlation reduces false alarms (e.g., HR + BP + clinical notes)
  - Experimental systems integrate bedside imaging (e.g., ultrasound + ventilator data)

MULTIMODAL APPLICATIONS IN PATIENT MONITORING



- Chronic Disease Management
  - **Diabetes:** Combine Continuos Glucose Monitor, fitness tracker (activity/HR), dietary logs, and insulin pump data
    - AI coaches detect lifestyle-related glycemic patterns
  - **Heart failure:** Integrate implantables (defibrillators, fluid sensors) with daily vitals and symptom diaries
    - Predictive models improve with fusion of RPM + EHR + prior labs
    - Multimodal detection of decompensation (e.g., ↑ weight + ↓ steps + ↑ resting HR)

- Wearable + Smartphone Multimodality
  - Smartphones collect:
    - Accelerometer (activity, gait), microphone (voice, cough), touchscreen (cognition/tremor), GPS
  - Used with wearables & self-reports to monitor:
    - Parkinson's (tremor + voice + med logs)
    - Mental health (mood, usage patterns, speech tone, activity)
- Real-World Examples

  - Apple Watch + Health app:
    - Tracks HR, blood oxigen, sleep, exercise + integrates EHR via Health Records
    - Aims to detect conditions using combined physiological + historical data
  - Philips HealthSuite, GE platforms:
    - Fuse hospital, wearable, and EHR data into cloud-based analytics systems
    - Building the infrastructure for scalable multimodal monitoring

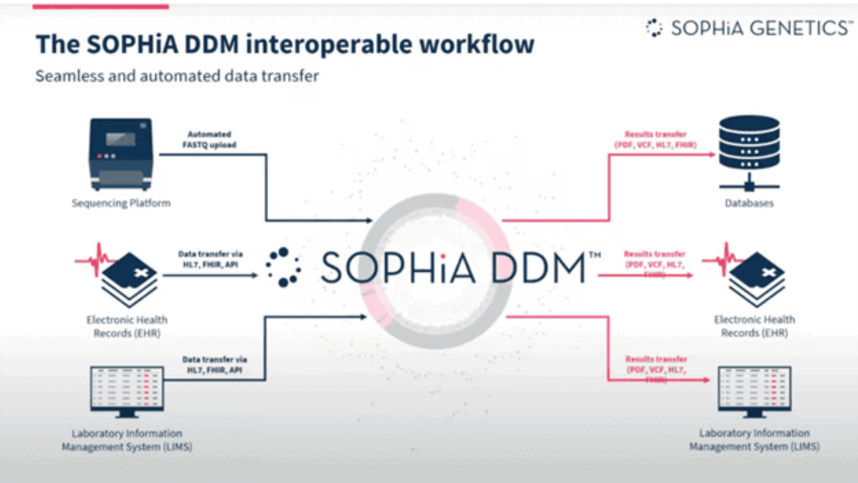
CURRENT SYSTEMS AND COMERCIAL EXAMPLES

IBM Watson

TEMPUS

aidoc aiOS™

biofourmis



https://medgemma.org/

MedGemma


[1, 5, 8]
CURRENT STATE OF THE FIELD: WHERE WE ARE

MULTIMODAL AI IN HEALTHCARE: PERFORMANCE & PROGRESS

- Improved Model Accuracy
  - Multimodal models consistently outperform single-modality ones (notable AUC/accuracy gains).
  - Enable predictions previously infeasible, e.g., long-term cognitive decline or treatment response.
- Enhanced Interpretability
  - Models generate rich outputs: annotated reports, visual-text links (e.g., phrase <-> radiology region).
  - Aids clinical decision-making and trust in AI outputs.
- State-of-the-Art Benchmarks
  - Top performance in medical VQA (image+text), pathology+genomics, and other complex tasks.
- Expanding Data Resources
  - Large datasets:
    - MIMIC-IV (ICU EHR + imaging + waveforms)
    - UK Biobank (500K+ with genomics, imaging, EHR, lifestyle)
    - ADNI, TCGA (disease-specific cohorts)
- Research Momentum
  - Surge in multimodal publications and challenges (e.g., *Nature Medicine* 2022 review).
  - Broad interest across academia and industry driving rapid advancement.

MULTIMODAL AI IN CLINICAL PRACTICE: ADOPTION & BARRIERS

- Early Clinical Integrations
  - Radiology AI (e.g., Aidoc): Combines imaging + clinical data for triage.
  - EHR-based Predictive Models: Flag high-risk patients (e.g., sepsis) using labs + notes.
  - Remote Monitoring: Wearables + EHR for chronic disease management.
  - Key Insight: Adoption occurs when AI clearly supports urgent clinical decisions.
- Limited Deployment in Routine Care
  - Few FDA-cleared multimodal products to date [8].
  - Translation Gap: Research often lacks deployment pathways.
  - Data Silos: Imaging (PACS), notes (EHR), labs stored separately → hinders fusion.
  - Trust & Transparency Issues: Clinicians resist black-box models lacking clear rationale.
  - Regulatory Hurdles: Complex approval for multi-source AI vs. single-modality tools.

CHALLENGES IN MULTIMODAL AI FOR HEALTHCARE

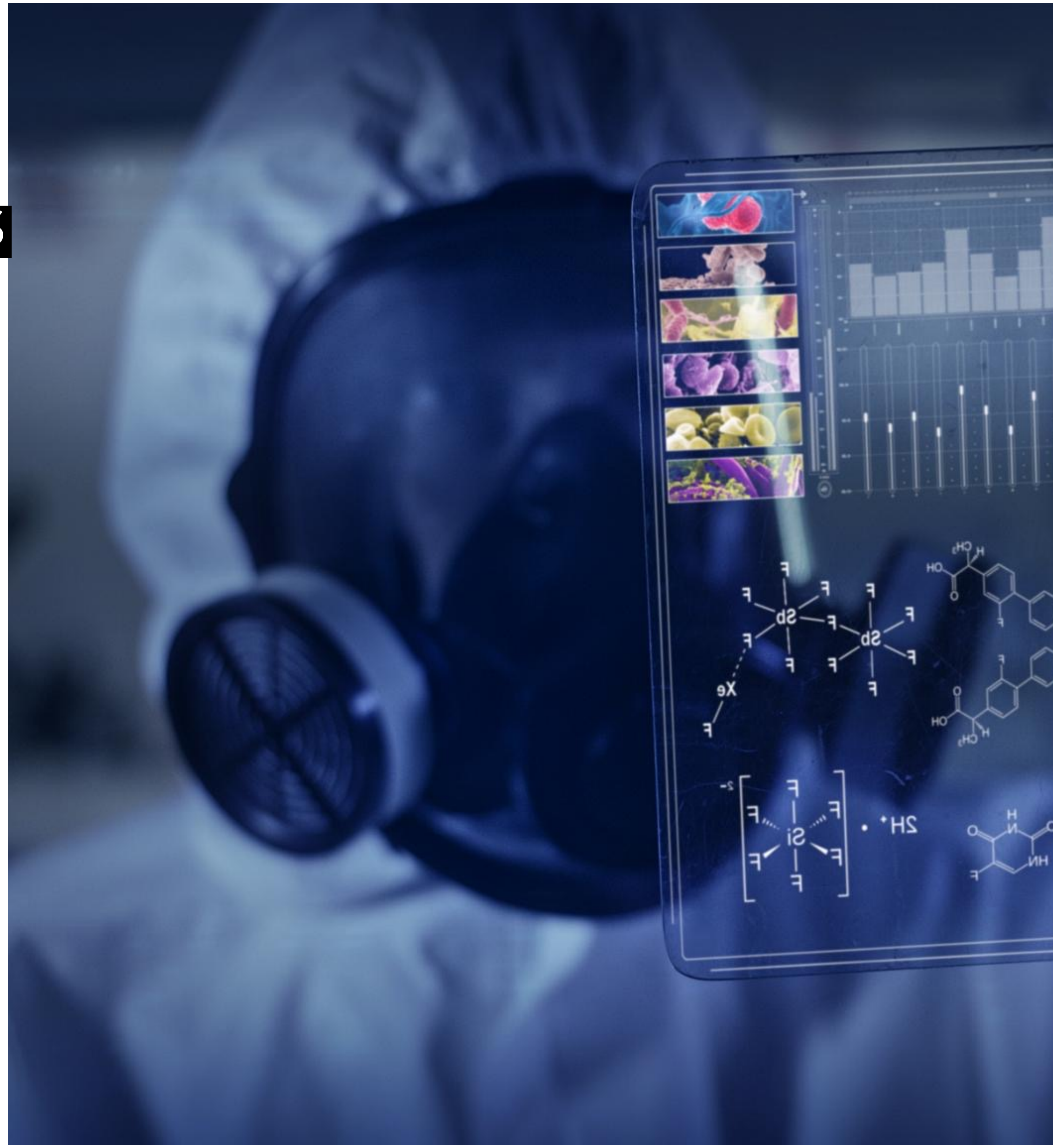
- Data Quality & Missing Modalities
  - Many patients lack complete data across all modalities (e.g., no MRI or genomics).
  - Models must handle incomplete inputs robustly.
- Bias Across Modalities
  - Imaging may underrepresent demographics; text may reflect clinician bias.
  - **Risk:** Biases can **compound** when modalities are combined.
- Computational Burden
  - Multimodal models require high compute (e.g., large images + long text + genomics).
  - **Deployment barriers:** Many hospital IT systems lack needed infrastructure.
- Lack of Standardization
  - Models are often **custom-built**, making comparisons and reproducibility hard.
  - No consensus yet on optimal fusion strategies (early vs. late fusion, etc.).
- Emerging Solutions
  - Community benchmarks (e.g., *Multimodal Hospital Mortality* on PhysioNet), but fragmented in methodology.
  - Convergence toward **transformer** and **graph-based** multimodal architectures. [5]

[1, 4, 5]

FUTURE TRENDS AND DIRECTIONS

MULTIMODAL FOUNDATION MODELS AND LARGE MODELS

- **Trend**: Rise of large pretrained models for multiple data types (text, images, etc.), inspired by GPT-4 and CLIP.
- **Healthcare Shift**: Development of **Multimodal Large Language Models (MLLMs)** trained on combined medical data (e.g., images + text).
- **Key Applications**:
  - **Medical VQA**: Answering natural language questions using radiology images.
  - **Image-to-Report**: Generating medical reports directly from images.
- **Training Approach**: Pretraining on large paired datasets (e.g., X-rays + reports) to learn cross-modal representations.
- **Examples**: Early models like **BioViL** and **MedCLIP** show strong generalization across tasks.
- **Vision**: Future "**AI residents**" — multimodal assistants that interpret both patient records and visuals to support clinicians.
- **Implication**: Foundation for intelligent hospital chatbots with access to full patient context (text + imaging).



IMPROVED DATA FUSION TECHNIQUES

- **From Concatenation to Intelligence**: Moving beyond simple feature concatenation toward **dynamic fusion** of modalities.
- **Advanced Architectures**:
  - **Cross-modal attention**: Highlights relevant info in one modality based on cues from another (e.g., image focus guided by clinical notes).
  - **Hierarchical models**: Extract modality-specific insights first, then combine (e.g., "enlarged heart" + "elevated troponin").
- **Improved Reasoning & Interpretability**: Mimics human diagnostic steps, supporting **transparent, stepwise reasoning**.
- **Use of Graph Neural Networks (GNNs)**:
  - Builds **patient-specific knowledge graphs** with diverse nodes (e.g., symptoms, genes, image findings).
  - Enables flexible inference and robust handling of **missing or evolving data**.
- **Future Direction**: Smarter, more adaptive fusion that reflects the **clinical context** and **task-specific needs**.

P4 MEDICINE: MULTIMODAL AI AS A CATALYST

- P4 Paradigm:
  Predictive, **Preventive**, **Personalized**, and **Participatory** healthcare.
- **Predictive**:
  Combines genomics, lifestyle, and imaging to forecast disease risk (e.g., diabetes).
- **Preventive**:
  Early detection through integrated monitoring (e.g., wearables spotting health decline).
- **Personalized**:
  Tailored care based on individual genetic, environmental, and clinical profiles.
- **Participatory**:
  Patients engage via apps/wearables; systems give direct feedback (e.g., nudges from smartwatch + EHR + genetics).
- **Beyond Prediction**:
  Future focus on **multimodal interventions** (e.g., adaptive coaching driven by multiple data streams).

SOLVING DATA SILOS FOR SCALABLE MULTIMODAL AI

- **Challenge**: Healthcare data remains siloed across systems and modalities.
- **Solution Path**:
  - **Adoption of standards**: FHIR, Fast Healthcare Interoperability Resources (clinical), DICOM, Digital Imaging and Communications in Medicine (imaging), etc.
  - **Creation of hospital data lakes**: Secure, unified repositories per patient.
  - **Unified APIs**: Enable access to EHR, imaging, and genomics together.
- **Infrastructure Investments**: Governments and health systems are building interoperable platforms.
- **Labeling Bottleneck**:
  - **Limited labeled multimodal data** slows progress.
  - Future relies on **weakly or self-supervised learning** to leverage unlabeled pairs (e.g., image-report or sensor-EHR alignments).
- **Outcome**: Greater scalability and performance by unlocking **vast, underutilized multimodal data**.

TRANSPARENCY, TRUST AND EXPLAINABILITY

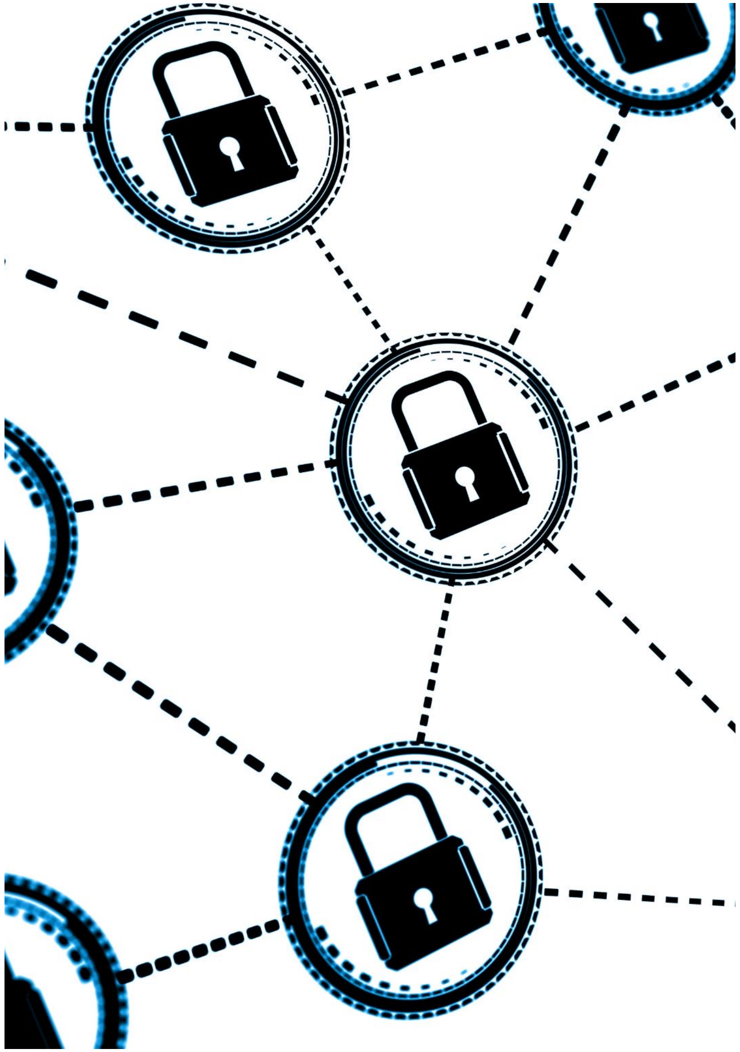
- **Challenge**: Multimodal models are harder to interpret than single-modality ones.
- **Key Strategies**:
  - **Interpretable outputs**: Highlight image regions + quote relevant clinical text to justify predictions.
  - **Causal modeling**: Explicitly model cause-effect links (e.g., gene → image finding) for transparency.
- **Building Trust**:
  - **Robust validation** across diverse settings and incomplete/noisy data scenarios.
  - **Explainability interfaces** to show what influenced decisions.
- **Regulatory Outlook**:
  - Future systems may be **required** to reveal which modalities drove a recommendation (e.g., image vs text vs genetics).
- **Goal**: Boost clinician confidence through **transparent, accountable AI tools**.

BIAS AND FAIRNESS CONSIDERATIONS

- **Risk**: Multimodal systems may **amplify existing biases** (e.g., unequal data availability across groups).
- **Examples**:
  - Imaging less available for rural or minority patients → model skews toward biased sources.
- **Research Priorities**:
  - **Bias detection** and **mitigation strategies**:
    - Balance training data across subpopulations.
    - Use adversarial learning to remove dependency on protected attributes.
- **Hopeful Outlook**:
  - If used correctly, **multiple modalities may reduce bias** by cross-validating inconsistent or biased inputs.
- **Key Need**: **Empirical validation** to ensure fairness gains are real and not just theoretical.

OPERATIONAL AND INTEGRATION CHALLENGES

- **Workflow Fit**:
  - Models must align with clinical routines (e.g., alerts, dashboards, timing of use).
  - **Implementation science** is key to making AI tools usable by care teams.
- **Financial Barriers**:
  - Few current reimbursement codes for AI tools.
  - **Incentives may emerge** as evidence grows (e.g., preventing readmissions).
- **Technical Robustness**:
  - Systems must handle **pipeline failures** (e.g., missing images, delayed labs).
- **User-Centered Design**:
  - Interfaces must deliver **multimodal insights without cognitive overload**.
  - **Human factors engineering** will be essential for clinician adoption.

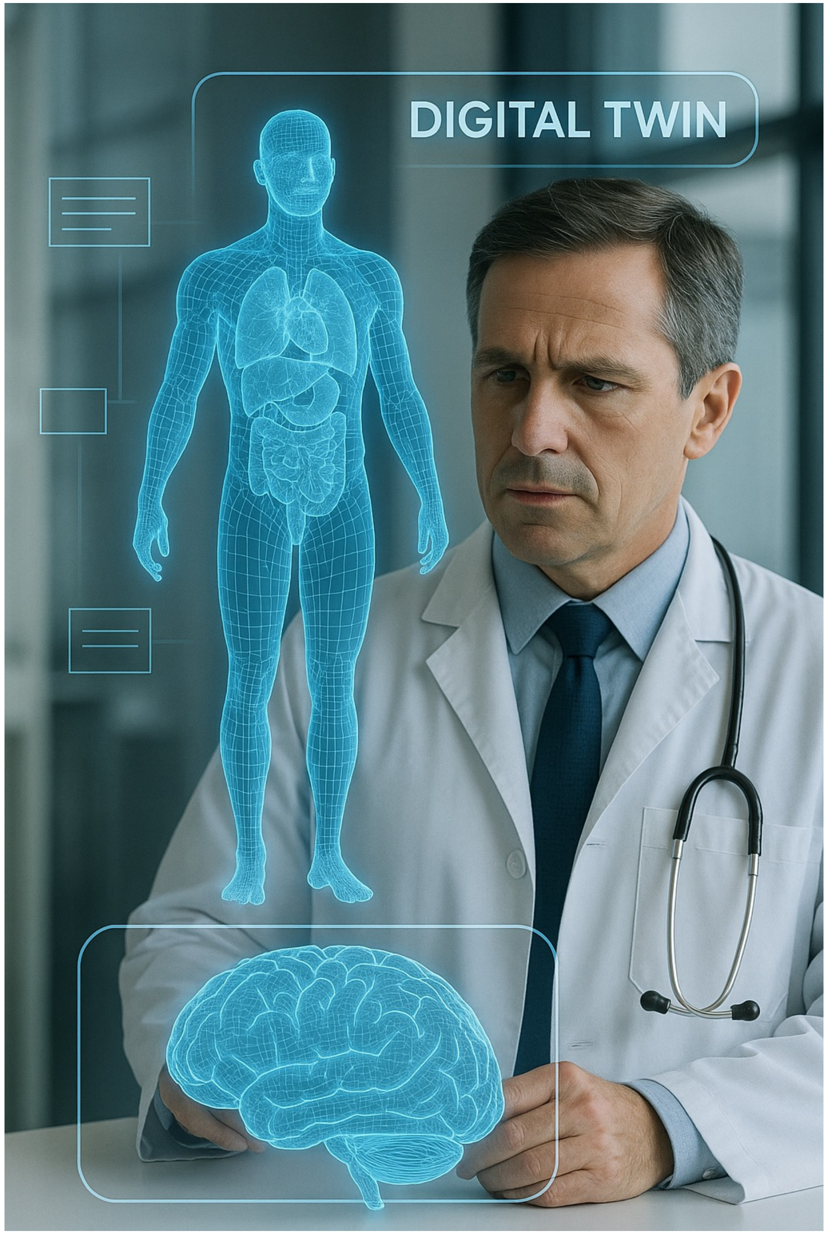


PRIVACY AND SECURITY

- Increased Risk:
  - **Data fusion** raises re-identification risk and impact of breaches.
  - More modalities = more **sensitive, comprehensive data exposed**.
- Privacy-Preserving Approaches:
  - **Federated learning** and **secure multi-party computation** allow cross-institutional training without centralizing data.
  - **Differential privacy** can prevent leakage when publishing models.
- Security Threats:
  - **Adversarial attacks** could target a single modality (e.g., tampered medical images).
  - Future systems must ensure **robustness across all modalities**.
- Key Priority: Build **secure, privacy-respecting infrastructures** for safe AI deployment.

EMERGING MODALITIES AND OPPORTUNITIES

- Growing Data Sources:
  - Beyond genomics: includes **microbiome**, **digital pathology**, and **pedigree (family history)** data.
- Digital Twins:
  - **Computational replicas** of patients integrating behavior, physiology, genes, etc.
  - Enables **virtual simulations of treatments** to predict optimal outcomes.
- Multimodal Virtual Health Assistants:
  - Combine **text/voice interaction** with real-time health data (e.g., wearables, records).
  - Support **personalized coaching** and **chronic disease management** outside clinics.
- Key Insight: The definition of "multimodal" will **keep evolving** as new health data streams emerge.



CONCLUSION

Advancement in Medical Support

Multi-modality approaches signify significant advancements in medical support systems, integrating various technologies.

Enhanced Diagnostics

Integrating diverse data sources enhances diagnostics, leading to more accurate and timely medical decisions.

Improving Patient Outcomes

By leveraging technology, we can significantly improve patient outcomes and overall healthcare quality.

Future of Personalized Medicine

The integration of these approaches paves the way for the future of personalized medicine tailored to individual needs.

REFERENCES

[1] Massimo Salvi, Hui Wen Loh, Silvia Seoni, Prabal Datta Barua, Salvador García, Filippo Molinari, U. Rajendra Acharya. Multi-modality approaches for medical support systems: A systematic review of the last decade. Inf. Fusion 103: 102134 (2024)

[2] Shaik, T. et al. (2023). "A Survey of Multimodal Information Fusion for Smart Healthcare: From Data to Wisdom." arXiv preprint

[3] Medigy/Forbes Tech Council (2025). "How Multimodal AI Is Impacting Healthcare." (Athira K.)

[4] Fountzilas, E. et al. (2025). "Convergence of AI/ML techniques in precision oncology." NPJ Digital Medicine, 8:75

[5] Has Multimodal Learning Delivered Universal Intelligence in Healthcare? A Comprehensive Survey https://arxiv.org/html/2408.12880v1

[6] Kline, A., Wang, H., Li, Y. *et al.* Multimodal machine learning in precision health: A scoping review. *npj Digit. Med.* **5**, 171 (2022).

[7] Mataraso, S.J., Espinosa, C.A., Seong, D. *et al.* A machine learning approach to leveraging electronic health records for enhanced omics analysis. *Nat Mach Intell* **7**, 293–306 (2025)

[8] Acosta, J.N., Falcone, G.J., Rajpurkar, P. *et al.* Multimodal biomedical AI. *Nat Med* **28**, 1773–1784 (2022).

[9] R. Mokni, N. Gargouri, A. Damak, D. Sellami, W. Feki, Z. Mnif, An automatic computer-aided diagnosis system based on the multimodal fusion of breast cancer (MF-CAD, Biomed. Signal Process Control. 69 (2021)

MULTI-MODALITY APPROACHES FOR MEDICAL SUPPORT SYSTEMS: WHERE WE ARE AND WHERE WE ARE GOING

Salvador García <salvagl@decsai.ugr.es>

Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI)

University of Granada